View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by UUM Repository

An Application of Network Topology to Understand The Signal in Process Variability: A Case Study in Petrochemical Industry

S. Sharif [1,2], M.A. Djauhari [1]

[1]Faculty of Science, Universiti Teknologi Malaysia, Malaysia

[2]College Art and Science, Universiti Utara Malaysia, Malaysia

(shamshurita@uum.edu.my, maman@utm.my)

Abstract - **This paper deals with a network topology approach for identifying the root causes of an out-of-control signal. An experience in the production process of fertilizer will be reported.**

Keywords - **centrality measure, correlation matrix, minimum spanning tree, out-of-control signal, production process control**

I. INTRODUCTION

Petrochemicals can be converted into thousands of industrial and consumer products including plastics, paints, rubbers, fertilizers, detergents, dyes, textiles and solvents. The industry consists of two major divisions. The primary petrochemical industry produces basic chemicals such as ethylene from oil or gas while the secondary industry converts the basic petrochemicals into materials that may be directly used by other industries. Most petrochemicals contain hydrogen or carbon or both.

This industry and the products play an enormous role in our daily lives, especially the fertilizer which is of our concern in this paper. Chemical fertilizers contain one or more of the essential growth nutrients such as nitrogen, phosphorus, and potassium and various others [1] . Once added to the soil, these nutrients fulfil the required demands of the plants. Due to the contribution of fertilizer towards industry, it is then important to control their quality of production. In our case study, five interrelated variables which determine the quality of fertilizer have been studied. Thus, multivariate statistical process control approach is needed [2] [3]

In this paper our discussion is focused on process variability control since managing and reducing variability is the primary concern in all quality improvement initiatives [4].

Furthermore, our discussion will be limited on the case where monitoring operation is based on subgroup observations. The remaining of the paper is organized as follows. In the next section we explain how to identify the root causes of an out-of-control (OOC) signal. In Section 3, we discuss the use of network topology approach to analyze the shift of correlation structure when an OOC signal occurs. Then, we present a case study on petrochemical industry in Section 4. At the end of this paper, we will draw attention to a conclusion.

II. UNDERSTANDING OOC SIGNAL

The generalized variance (GV) chart [5], [6], [7], [8], and vector variance (VV) chart [9]. [10] are used in this paper to monitor a multivariate process variability based on subgroup observations. When the GV chart fails to detect the shift of covariance structure, the VV chart might be able to do the job and vice versa. Thus, if VV and GV charts are simultaneously used, they will provide a much better understanding of the multivariate variability [11].

Once an out-of-control (OOC) signal occurs in either GV chart or VV chart during a process variability monitoring operation, we have to be able to explain why such signal occurs.

In process mean vector monitoring, if an OOC signal occurs, the MYT decomposition method [11] can be used to analyze the root causes of the signal. However, in process variability monitoring there is no special method except in the case of individual observations [12]. In the case of subgroup observations, the interpretation of OOC signal is not an easy task. The complexity of covariance structure is the source of the problem. Therefore, in order to have a better understanding of covariance structure, multivariate methods such as, among others, principal component analysis (PCA) and clustering analysis [13] can be used to simplify the analysis of the structure.

In this paper, we used an approach based on network topology developed in the field of econophysics [14].

We propose the following procedure to analyze the root causes of an OOC signal in terms of the shift of correlation structure.

Step 1: Construct control chart and use it to monitor the process variability. If an OOC signal occurs then continue to Step 2. Otherwise, further action is not required and monitoring operation is continued.

Step 2: Construct the network topology of the current process variability and the reference sample variability.

Step 3: Compute the centrality measures of both networks to understand the current situation.

Step 4: Determine the variables having the highest scores of those measures.

978-1-4577-0739-1/11/$26.00 ©2011 IEEE

III. NETWORK TOPOLOGY

Network topology is one way to visualize a complex system represented by a dissimilarity matrix. It was originally developed in computer science and, as can be seen in the literature, has been used in various fields. The essence of network topology lies in its elements and the way they connect with each other [15].

In order to construct a network topology, first we construct the correlation matrix and then transform into to a distance matrix. Finally, from distance matrix we construct the corresponding minimum spanning tree (MST) which simplifies the original network in the form of tree by using Kruskal algorithm [16].

If there are p variables involved in the process, then we will have $p[p-1]/2$ elements of correlation coefficient to investigate. The role of MST is to filter the information contain in correlation structure into p-1 optimal correlation structure.

To interpret the MST, the centrality measures such as degree, betweeness, closeness and eigenvector centrality are usually used in practice. These measures will help to understand the importance role of each variable in the process relative to the others [17], [18], [19] [20] and [21].

The advantage of network topology as a tool to simplify the complex structure of correlation matrix is not only lies in its ability to visualize that structure but also in giving the interpretation of that structure which is useful in the search of root causes of any OOC signal. This will be clarified in the following case study.

IV. A CASE STUDY

Here we present a case study on production process of ZA fertilizer at carbonation level at a petrochemical industry. There are p = 5 variables that determine the quality of fertilizer namely, CO_2 (gr/lt), NH_3 (gr/lt), Ratio CO_2/NH_3, BD (kg/lt), and temperature (^{0}C). The data were collected during one month or 22 working days. There are two subgroups in each working day. Therefore the number of independent subgroups is m = 44. The subgroup size is n = 6.

In this section we use GV chart [7] and VV chart [9] to monitor the production process variability of fertilizer. To construct the GV chart, first we calculate the value of GV for each subgroup. Second, the determinant of the average of all covariance matrices, $|\overline{S}|$ = 0.0106. The result is presented in TABLE 1. Then, we determine the control limits as in [21],

$$\text{UCL} = |\overline{S}|\left(\frac{b_1}{b_3} - 3\sqrt{\frac{b_2}{b_3^2+b_4}}\right) \text{ and LCL} = |\overline{S}|\left(\frac{b_1}{b_3} + 3\sqrt{\frac{b_2}{b_3^2+b_4}}\right)$$

where b_1, b_2, b_3 and b_4 are calculated from data.

Since b_1 = 0.0384, b_2 = 0.0295 b_3 0.9553, and b_4 = 0.0426, then we obtain LCL = -0.0052 which is set to 0 and UCL = 0.006. The GV chart is presented in Fig. 1. It shows that all data points are under UCL, meaning that there is no OOC signal occurs.

TABLE 1
SAMPLE GV

Sample	GV	Sample	GV
1	3.42E-11	23	6.34E-13
2	3.84E-24	24	7.57E-35
3	4.26E-09	25	1.36E-08
4	1.52E-08	26	1.46E-07
5	0.00E+00	27	0
6	1.81E-09	28	7.73E-07
7	2.00E-11	29	1.30E-10
8	1.24E-09	30	2.09E-07
9	3.92E-06	31	2.14E-11
10	7.28E-09	32	4.62E-08
11	5.81E-12	33	3.63E-08
12	1.09E-23	34	0
13	3.52E-09	35	3.22E-10
14	2.46E-10	36	1.01E-09
15	4.99E-05	37	9.23E-12
16	0	38	1.93E-09
17	1.33E-08	39	0
18	2.11E-34	40	2.48E-11
19	4.21E-22	41	2.20E-09
20	1.49E-06	42	1.19E-10
21	2.62E-09	43	0
22	1.13E-09	44	1.88E-11

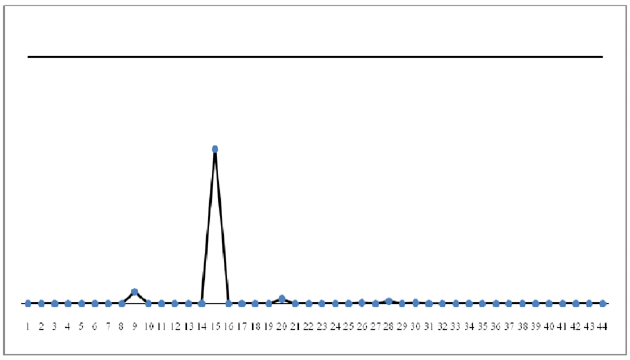

Fig. 1. GV chart

To construct VV chart, we need to calculate VV i.e., the trace of the squared covariance matrix for each subgroup. The result is presented in TABLE 2.

TABLE 2
SAMPLE VV

Sample	VV	Sample	VV
1	346	23	7214
2	1785	24	34166
3	191799	25	14988
4	329006	26	19621
5	4588	27	5261
6	1739	28	274
7	2518	29	118009
8	2006	30	455114
9	25668	31	697578
10	64913	32	189659
11	31875	33	7545
12	10694	34	2976
13	5635.85	35	39324
14	21523	36	9484
15	2456582	37	10812
16	594	38	30409
17	39295	39	5589
18	55364	40	14061
19	1542	41	18051
20	87060	42	9307
21	10007	43	22003
22	2693	44	4412

From the data set, following the formula in [9], we compute the squared covariance matrix, the sum of all its diagonal elements, and the sum of square of all its elements. We get $Tr\left(\bar{S}^2\right)$ = 26138 and $Tr\left(\bar{S}^4\right)$ = 431984850. From these results, $\hat{\theta}$ = 3.6264e+04 and $\hat{\eta}$ = 1.3032e+09. See [22] for the formula of $\hat{\theta}$ and $\hat{\eta}$. Therefore, *LCL*= $\hat{\theta}-3\frac{\hat{\eta}}{\sqrt{n-1}}$ = - 1.748e+09 which is set to 0 since it is negative, and *UCL*= $\hat{\theta}+3\frac{\hat{\eta}}{\sqrt{n-1}}$. = 1.748e+09. The observation values of VV together with *LCL* and *UCL* are visualized in the so-called VV chart in Fig. 2.

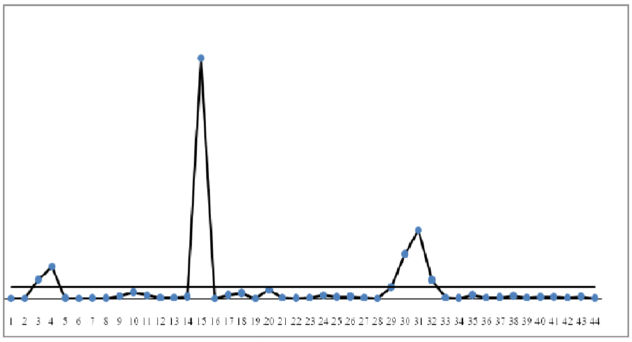

Fig. 2. VV chart

In this figure we can see that, according to VV chart, six OOC signals occur at the third, fourth, fifteenth, thirtieth, thirty first, and thirty second samples. To identify the root causes of each signal, in what follows we only consider the fifteenth sample as an example because it is the most dominance signal. The other signal can be analyze in the same manner. To use network topology approach, first we construct the correlation matrix among the five variables at the fifteenth sample. The result is in TABLE 3 compare it with the correlation matrix in reference sample in TABLE 4.

TABLE 3
CORRELATION MATRIX OF FIFTEENTH SAMPLE

	CO_2	NH_3	Ratio	BD	Temp
CO_2	1				
NH_3	0.091	1			
Ratio	0.721	0.265	1		
BD	0.678	0.416	0.739	1	
Temp	-0.097	-0.169	-0.330	-0.447	1

TABLE 4
CORRELATION MATRIX OF REFERENCE SAMPLE

	CO_2	NH_3	Ratio	BD	Temp
CO_2	1				
NH_3	0.246	1			
Ratio	0.648	-0.395	1		
BD	0.134	0.004	0.113	1	
Temp	0.230	-0.011	0.1958	0.009	1

Next, we transform the correlation matrix C into a distance matrix D by using $d_{ij}=\sqrt{2(1-c_{ij})}$ [14]. The higher the correlation in [-1, 1], the smaller the distance in [0,2], and the lower the correlation, the larger the distance. TABLE 5 and TABLE 6 represent the distance matrix related to fifteenth sample and to reference sample, respectively.

TABLE 5
DISTANCE MATRIX OF FIFTEENTH SAMPLE

	CO_2	NH_3	Ratio	BD	Temp
CO_2	1				
NH_3	1.35	1			
Ratio	0.75	1.21	1		
BD	0.80	1.08	0.72	1	
Temp	1.48	1.53	1.63	1.70	1

TABLE 6
DISTANCE MATRIX OF REFERENCE SAMPLE

	CO_2	NH_3	Ratio	BD	Temp
CO_2	1				
NH_3	1.23	1			
Ratio	0.84	1.67	1		
BD	1.32	1.41	1.33	1	
Temp	1.24	1.42	1.27	1.41	1

By using *Pajek* software [15], we represent TABLE 5 and TABLE 6 in the form of an indirected weighted complete graphs in Fig. 3 and Fig. 4, respectively. These graphs show the degree of interrelationship among variables. Their corresponding MST in Fig. 5 and Fig. 6.

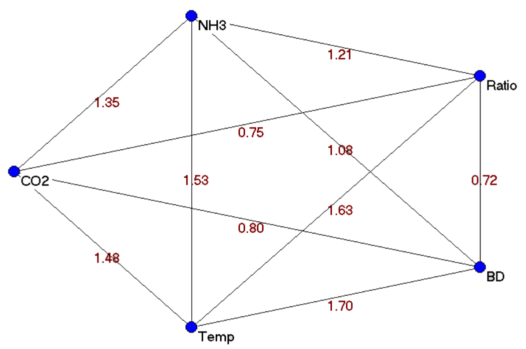

Fig. 3. The relationship among five quality characteristics of fifteenth sample

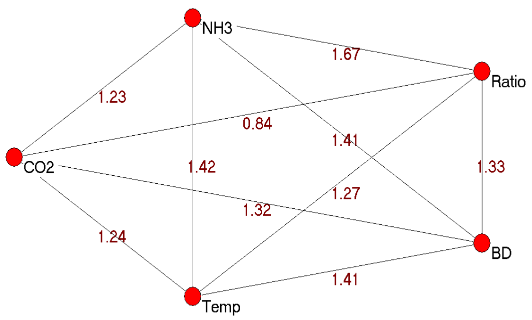

Fig. 4. The relationship among five quality characteristics of reference sample

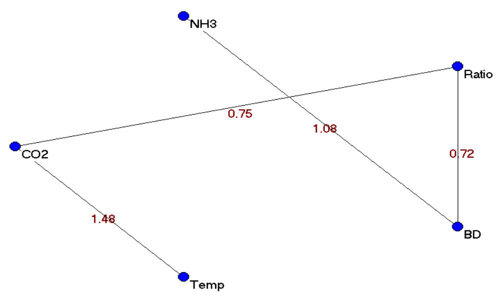

Fig 5. MST of fifteenth sample

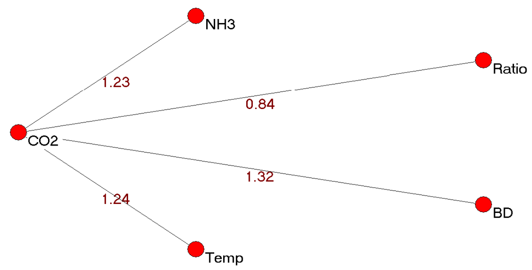

Fig 6. MST of reference

To summarize the information contained in those MSTs, in TABLE 7 and TABLE 8 we present the score of degree, betweeness, closeness, and eigenvector centrality measures.

TABLE 7
CENTRALITY MEASURE OF FIFTEENTH SAMPLE

Quality Variables	Degree	Betweenness	Closeness	Eigenvector
CO_2	0.50	0.50	0.57	0.50
NH_3	0.25	0	0.40	0.288
Ratio	0.50	0.67	0.67	0.577
BD	0.50	0.50	0.57	0.5
Temp	0.25	0	0.40	0.288

TABLE 8
CENTRALITY MEASURE OF REFERENCE SAMPLE

Quality Variables	Degree	Betweenness	Closeness	Eigenvector
CO_2	1	1	1	0.707
NH_3	0.25	0	0.571	0.353
Ratio	0.25	0	0.571	0.353
BD	0.25	0	0.571	0.353
Temp	0.25	0	0.571	0.353

From TABLE 7 and TABLE 8, we learn that:

(i) CO_2, Ratio and BD have the highest number of connections (2) in the network based on degree centrality. Each of the followings has only one (1) connection: NH_3 and Temp. The higher the number of connections the more influential of a particular variable.

(ii) In terms of betweenness, Ratio plays the most important role in the network. This means that if those characteristics are well managed, then the others will be influenced.

(iii) Regarding the closeness, Ratio has an excellent position compared to the others where the information flow in the network can easily reach others.

(iv) According to eigenvector centrality, Ratio act as the most influential variable to the others in this network. They have the highest degree to which it has formed strong relationship with other four influential characteristics, ordered decreasingly in terms of eigenvector score: BD and. CO_2, NH_3, and Temp.

In summary, according to the four (4) centrality measures, the following variables occur at least for one measure: CO_2, BD and Ratio. The highest number of occurrence is for Ratio (4 times).

V. CONCLUSION

In this paper we show how to identify an out-of-control signal in fertilizer production process using network topology. Five quality variables and their correlation structure are considered as a network system. To simplify that network we used minimum spanning tree which provide an optimal sub-network in the form of a spanning tree. This tree is then used to construct the optimal network topology of those variables. To summarize the information contain in that network topology, the four centrality measure such as degree centrality, betweeness centrality, closeness centrality, and eigenvector centrality are used.

According to the above analysis, among five quality variables, Ratio and followed by CO_2 are the most important variables that influence the occurrence OOC signal at the fifteenth sample.

ACKNOWLEDGMENT

This research is sponsored by the Ministry of Higher Education of Malaysia under vote numbers 78484 and 4F013 and Q.J13000.7126.02H18.. The authors gratefully acknowledge those sponsorships and Universiti Teknologi Malaysia for providing us with research facilities. They also thank the reviewers for their helpful comments and suggestions.

REFERENCES

[1] C.O. Plank. (1989) *Plant analysis handbook for Georgia. Athens (GA):* University of Georgia Cooperative Extension Service, [Online]. Available*:* http://aesl.ces.uga.edu/publications/plant/Nutrient.htm

[2] W. H. Woodall, and D.C. Montgomery. "Research issues and ideas in statistical process control", *Journal of Quality Technology*. vol. 31, no. 4, pp. 376-386, 1999.

[3] S.Bersimis, S. Psarakis, and J. Panaretos. "Multivariate Statistical Process Control Charts: An Overview", *Quality and Reliability Engineering International*, vol. 23, pp. 517-543, 2007.

[4] F. W. Breyfogle. *Implementing Six Sigma: Smarter Solution using Statistical Methods* , 2nd Ed, Hoboken, NJ: Wiley,. 2003.

[5] F. B. Alt, and N.D. Smith. *Multivariate Process Control,* Handbook of Statistics, vol. 7, pp. 333-351, (1988).

[6] D.C. Montgomery. *Introduction to statistical quality control*, 5th ed., New York : Wiley. 2004.

[7] M.A. Djauhari. "Improved Monitoring of Multivariate Process Variability". *Journal of Quality Technology*, vol.37 no.1, pp 32-39*,* 2005.

[8] R. L. Mason , Y. M. Chou, and J.C. Young.."Detection and Interpretation of a Multivariate Signal Using Combined Charts". *Communications in Statistics - Theory and Methods*, vol. 40, no.5, pp. 942-957, 2011.

[9] M. A. Djauhari, , M. Mashuri, and D. E. Herwindiati, "Multivariate process variability monitoring", *Communication in Statistics – Theory and Method*, vol.37, no.11, pp. 1754, 2008.

[10] R. L. Mason , Y. M. Chou, and J.C. Young. "Decomposition of Scatter Ratios Used in Monitoring Multivariate Process Variability". *Communications in Statistics-Theory and Methods*, vol.39. no.12, pp 2128-2145, 2010.

[11] R.L. Mason, and C.J. Young. *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: SIAM, 2002.

[12] R.L. Mason, Y.M. Chou and C.J. Young, "Decomposition of scatter ratios used in monitoring multivariate process variability," *Communication in Statistics – Theory and Methods*, vol 39, pp 2128-2145, 2010.

[13] R. A. Johnson and D. A. Wichern. *Applied Multivariate Statistical Analysis,* 5th Ed., Prentice Hall, 2*002*

[14] R.N. Mantegna, and H.E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge UK : Cambridge University Press, 2000.

[15] G. Bounova. "Topological Evolution of Networks : Case Studies in the US Airlines and Language Wikipedias; PhD Desertation. Massachusetts Institute of Technology, 2009.

[16] R. N. Mantegna, "Hierarchical Structure in Financial Markets", *European Physical Journal B* , vol. 11, pp.193-197, 1999.

[17] K. Park and A. Yilmaz. "A Social Network Analysis Approach to Analyze Road Networks," presented at the *ASPRS Annual Conference*, San Diego, CA. 2010.

[18] Y. Xu, , J. Ma, Y. Sun, J. Hao, Y. Sun and Y. Zhao, "Using Social Network Analysis As A Strategy For E-Commerce Recommendation", in *Proc. Pacific Asia Conference on Information Systems (PACIS). India.*

[19] A. Abbasi, and J. Altmann, "On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks," Technology Management, Economics and Policy Papers, Seoul National University no.2010, pp.66.

[20] J. Monárrez-Espino, and J. R. Caballero Hoyos, "Stability of Centrality Measures in Social Network Analyses to Identify Long-Lasting Leaders from an Indigenous Boarding School of Northern Mexico", *Estudios sobre las Culturas Contemporaneas*, vol. 16, no.32, pp. 155-171. , 2010.

[21] S.P. Borgatti, "Centrality and AIDS," *Connections*, vol 18, no.1, pp. 112-114, 1995.